PadhAI: Information Theory and Cross Entropy

One Fourth Labs

Relation to Number of Bits

Relation between number of bits and entropy

1. Consider the Entropy equation from the previous section using shorthand P_i for P(X=i)
2. $H(X) = -\Sigma_{i \in \{A,B,C,D\}} P_i * log\ P_i$
3. Suppose there is a message X that you want to transfer that can take 4 values: A, B, C, D
4. For 4 values, we would use 2 Bits to transfer each message

Random Variable: X	**2 Bit version**	**Probability Distribution: P(X=?)**	**Information Content: IC(X=?)**
A	00	1/4	$-log_2 2^2 = 2$ (ie $log_a a^n = n$)
B	01	1/4	$-log_2 2^2 = 2$
C	10	1/4	$-log_2 2^2 = 2$
D	11	1/4	$-log_2 2^2 = 2$

5. Now we can make the connection that the number of bits required to transfer a message is equal to the information content of that message
6. Consider another message X with 8 values: A, B, C, D, E, F, G, H

Random Variable: X	**3 Bit version**	**Probability Distribution: P(X=?)**	**Information Content: IC(X=?)**
A	000	1/8	$-log_2 2^3 = 3$ (ie $log_a a^n = n$)
B	001	1/8	$-log_2 2^3 = 3$
C	010	1/8	$-log_2 2^3 = 3$
D	100	1/8	$-log_2 2^3 = 3$
E	011	1/8	$-log_2 2^3 = 3$
F	101	1/8	$-log_2 2^3 = 3$
G	110	1/8	$-log_2 2^3 = 3$
H	111	1/8	$-log_2 2^3 = 3$

7. While sending a continuous stream of messages, we would be interested in minimizing the stream of bits that we send
8. Consider the same 4 valued example but with a different distribution

Random Variable: X	**Probability Distribution: P(X=?)**	**Information Content: IC(X=?)**
A	1/2 (High prob)	$-log_2 2^1 = 1$ (ie $log_a a^n = n$)
B	1/4 (Medium prob)	$-log_2 2^2 = 2$
C	1/8 (Low prob)	$-log_2 2^3 = 3$
D	1/8 (Low prob)	$-log_2 2^3 = 3$

9. This situation is considered favourable only if the average number of bits is less that the value it takes for an equally distributed set of values
10. The average is calculated using Entropy $H(X) = -\Sigma_{i \in \{A,B,C,D\}} P_i * log\, P_i$
11. Average/Entropy $= \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1.75\ which\ is\ < 2$
12. Thus, the Entropy gives us the ideal number of bits that should be used to transmit the message